Support Vector Machine (SVM) Documentation

Description: Classify samples using SVM
Author: Joshua Gould, gp-help@broad.mit.edu

Summary: The Support Vector Machines (SVM) module implements the support vector machines algorithm for supervised classification methods to compute a maximal separating hyperplane between the expression vectors of different classes or phenotypes. Given microarray data with *n* genes per sample, SVM outputs a hyperplane, w, which can be thought of as a vector with *n* components each corresponding to the expression of a particular gene. Loosely speaking, assuming that the expression values of each gene have similar ranges, the absolute magnitude of each element in w determines its importance in classifying a sample, since,

$$f(x) = \sum_{i=1}^{n} w_i x_i + b ,$$

and the class label is $\operatorname{sign}[f(x)]$.

The SVM model can be tested on a separately specified test set. Additionally, the model can be saved and used subsequently on additional test sets.

The table below summarizes the different options available and which parameters are required depending on the option selected.

Parameter	Train create a predictive model from a training dataset	Test with saved model run a saved model on a new test dataset	Train/Test create a model on training data and run it on test data
train.data,filename	Required	No	Required
train.cls.filename	Required	No	Required
saved.model.filename	No	Required	No
test.data.filename	No	Required	Required
test.cls.filename	No	Required	Required
pred.results.file	No	Yes	Yes
model.output.file	Required	No	Required

References:

* R. Rifkin, S. Mukherjee, P. Tamayo, S. Ramaswamy, C-H Yeang, M. Angelo, M. Reich, T. Poggio, E.S. Lander, T.R. Golub, J.P. Mesirov, *An Analytical Method for Multiclass Molecular Cancer Classification*, SIAM Review, 45:4, (2003).

* T. Evgeniou, M. Pontil, T. Poggio, *Regularization networks and support vector machines*, Adv. Comput. Math., 13 (2000), pp. 1-50.

* V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

Parameters:

Name	Description
train.data.filename	The training data file - .gct, .res
train.cls.filename	The training class file - .cls
saved.model.filename	A saved model from a previous run of SVM
test.data.filename	The test data file - .gct, .res
test.cls.filename	The test class file - .cls
pred.results.file	The name of the output file for prediction results
model.output.file	The name of the file containing the saved model

Output Files:

        If test data is supplied:

                1. file containing the prediction results

        If training data is specified

                1. a file containing the saved prediction model

Platform dependencies:

        Module type: Prediction

        CPU type: any

        Language: R